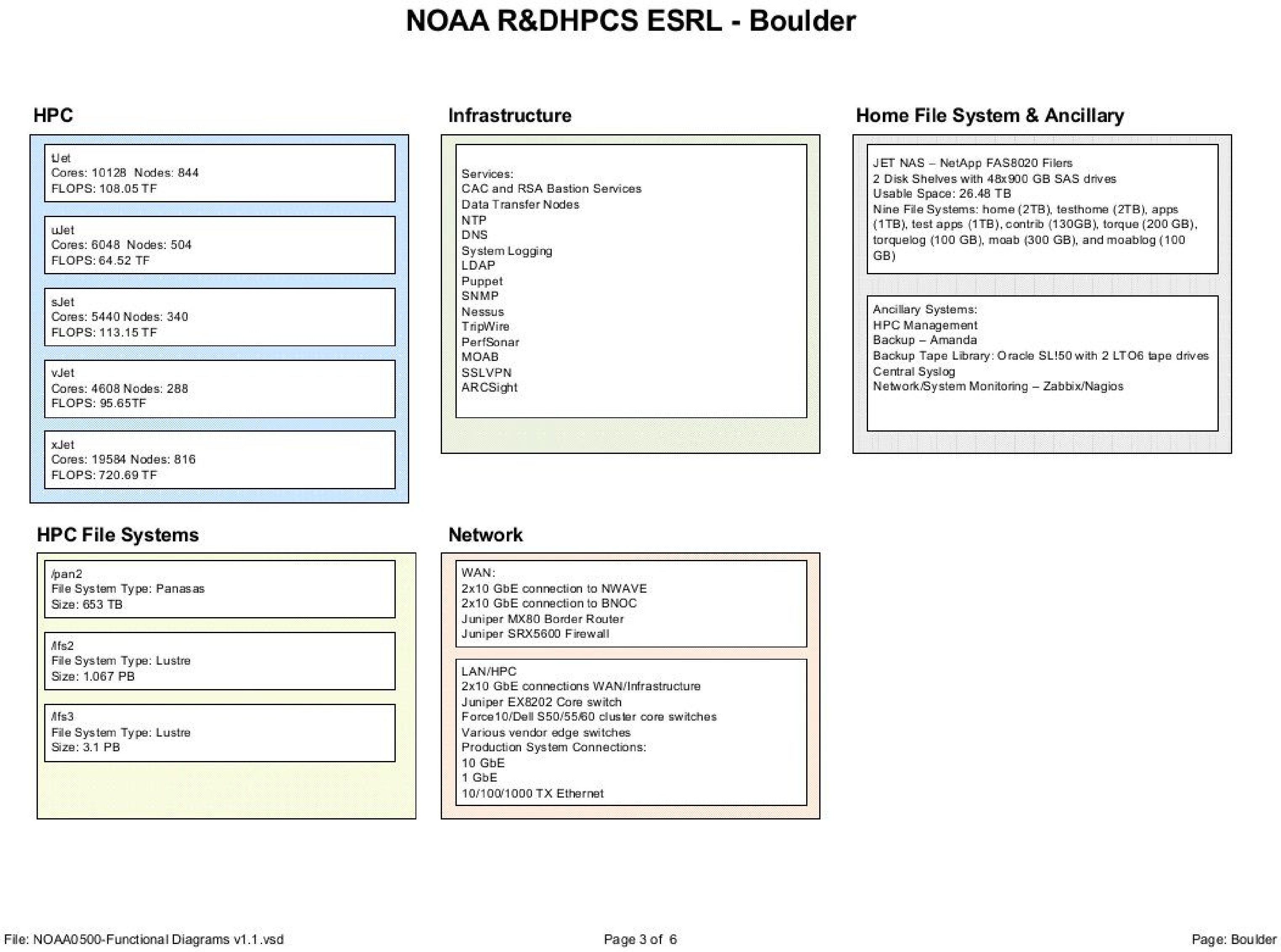
NOAA R&D HPCS RFP

Appendix A.1

~



NOAA R&DHPCS ESRL - Boulder

HPC

tJet
Cores: 10128 Nodes: 844
FLOPS: 108.05 TF

uJet
Cores: 6048 Nodes: 504
FLOPS: 64.52 TF

sJet
Cores: 5440 Nodes: 340
FLOPS: 113.15 TF

vJet
Cores: 4608 Nodes: 288
FLOPS: 95.65TF

xJet
Cores: 19584 Nodes: 816
FLOPS: 720.69 TF

Infrastructure

Services:
CAC and RSA Bastion Services
Data Transfer Nodes
NTP
DNS
System Logging
LDAP
Puppet
SNMP
Nessus
TripWire
PerfSonar
MOAB
SSLVPN
ARCSight

Home File System & Ancillary

JET NAS – NetApp FAS8020 Filers
2 Disk Shelves with 48x900 GB SAS drives
Usable Space: 26.48 TB
Nine File Systems: home (2TB), testhome (2TB), apps (1TB), test apps (1TB), contrib (130GB), torque (200 GB), torquelog (100 GB), moab (300 GB), and moablog (100 GB)

Ancillary Systems:
HPC Management
Backup – Amanda
Backup Tape Library: Oracle SL!50 with 2 LTO6 tape drives
Central Syslog
Network/System Monitoring – Zabbix/Nagios

HPC File Systems

/pan2
File System Type: Panasas
Size: 653 TB

/lfs2
File System Type: Lustre
Size: 1.067 PB

/lfs3
File System Type: Lustre
Size: 3.1 PB

Network

WAN:
2x10 GbE connection to NWAVE
2x10 GbE connection to BNOC
Juniper MX80 Border Router
Juniper SRX5600 Firewall

LAN/HPC
2x10 GbE connections WAN/Infrastructure
Juniper EX8202 Core switch
Force10/Dell S50/55/60 cluster core switches
Various vendor edge switches
Production System Connections:
10 GbE
1 GbE
10/100/1000 TX Ethernet

File: NOAA0500-Functional Diagrams v1.1.vsd Page 3 of 6 Page: Boulder

Detailed HPCS Architecture Boulder

16 January 2017

Boulder Facility Organization and Subsystems Overview

The Boulder HPC systems are divided amongst two rooms in the David Skaggs Research Center (DSRC) in Boulder, CO. Room 2B407 is primarily for low power density air cooled equipment, such as parallel file systems, Front Ends/Login nodes, Batch nodes, Infrastructure servers, and Home File System. There is also a 288 node compute cluster located in 2B407. However, this cluster requires a special air curtain for cold air isolation.

Room GA405 is for high density air cooled equipment such as compute racks. Each compute rack is comprised of 60-70 nodes, IB switches and Ethernet switches. Four of the five compute clusters are located on the GA405 room. Both rooms are connected via 18 active, 150 meter, QDR IB links. So all compute systems reside on the same IB fabric.

Jet is comprised of 5 separate clusters with a total core count of 45,808 cores. A single job cannot span multiple clusters. All clusters can be accessed through the Batch system from the common Front-ends/Login nodes.

System	Intel Processor	Install Date	Clock (GHz)	Nodes	Total Cores	Memory Per Node (GB)	Memory Per Core (GB)
tJet	Westmere	8/2010	2.66	758	9096	24	2
uJet	Westmere	11/2011	2.66	590	7080	24	2
sJet	Sandy Bridge	8/2012	2.6	340	5440	32	2
vJet	Ivy Bridge	8/2014	2.6	288	4608	64	4
xJet	Haswell	9/2015 8/2016*	2.3	816	19584	64	2.67

* xJet expansion

Jet has approximately 4.82PB of scratch storage. This is across 3 parallel file-systems (2 DDN Lustre and 1 Panasas). Combined peak performance of all three scratch file systems is ~46GB/s. All clusters reside on the same High Speed Interconnect/Infiniband fabric, making the parallel file systems globally accessible from all nodes. There is also a Home File System (HFS) which is globally accessible from all nodes over the Ethernet network.

The Home File System (HFS) and other NFS file systems are being served by a NetApp FAS8120. The HFS is globally accessible from all nodes over the Ethernet network.

The HFS and HPC infrastructure servers are backed up on a daily and weekly basis to a combination of spinning disk and tape. The Open Source software Amanda is used for all backups.

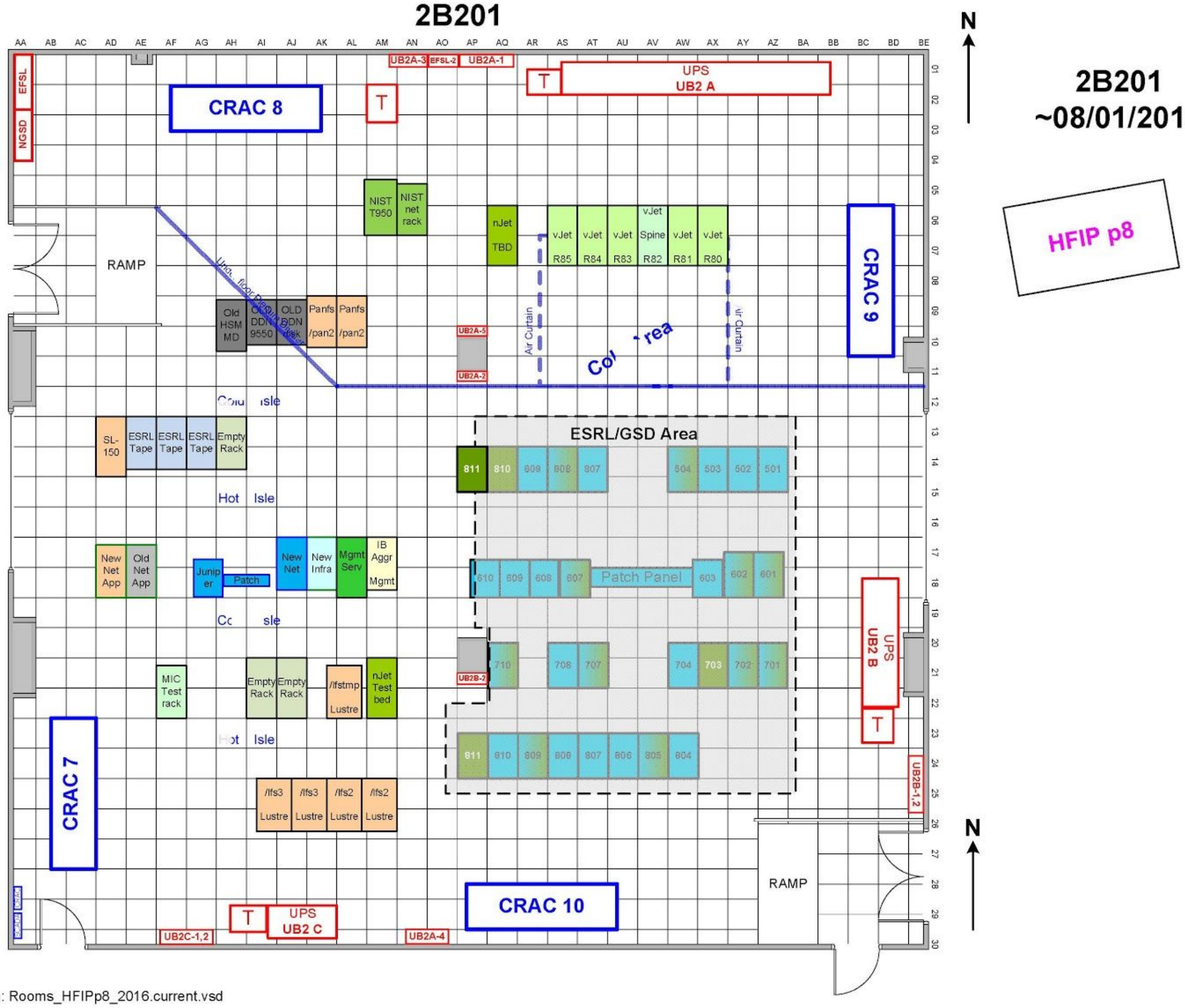
There is a tape storage archive located at the NESCC site in Fairmont, WV. Data can be transferred to and from this archive from Jet’s Front Ends. The retention period for the archive is 1-5 years or permanent. All projects with a compute allocation are allowed to store data on the archive.

Data transfer services are provided by a combination of 10GbE connected Data Transfer Nodes (DTN’s) and the 10GbE connected Front Ends. The DTN’s are used when a data transfer is initiated from outside of Jet. The Jet FE’s are used when a data transfer is initiated from Jet.

Jet’s Batch system is comprised of three separate pieces of software. Torque is the resource manager and is responsible for job submission, node health status and job launching. Moab is the scheduler and is responsible for scheduling jobs based on priority and managing dedicated system resources with reservations. MAM in the allocation manager and is responsible for managing the monthly allocation of core hours for each project. MAM and Moab are also used for the daily and monthly reporting of system utilization.

Diagrams

NOAA/Boulder



File: Rooms_HFIPp8_2016.current.vsd

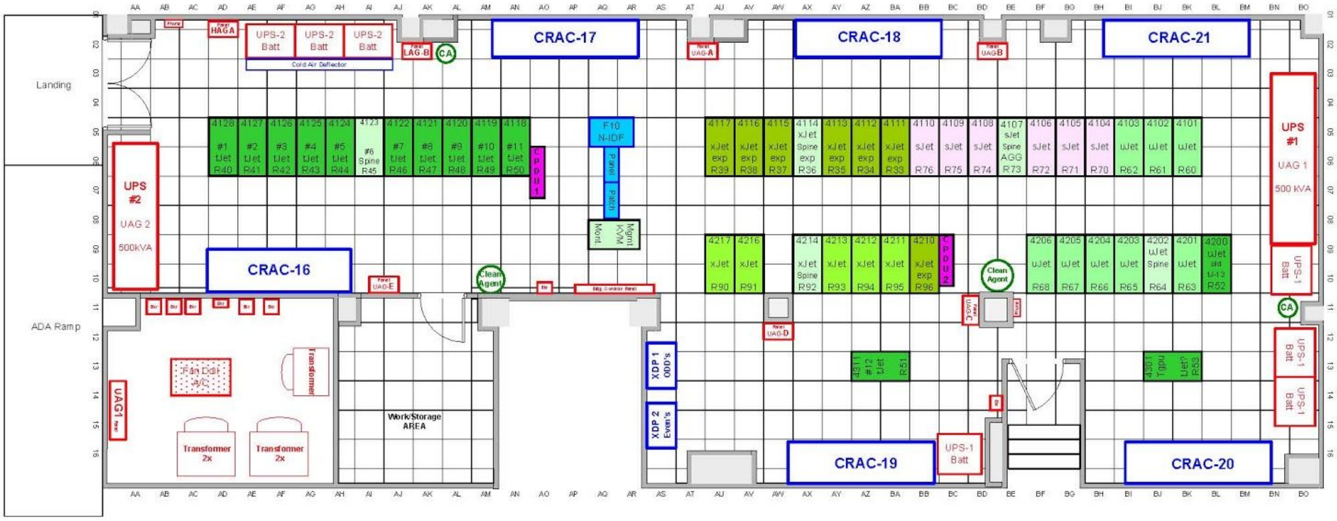
GA405
~08/01/2016

HFIP p8

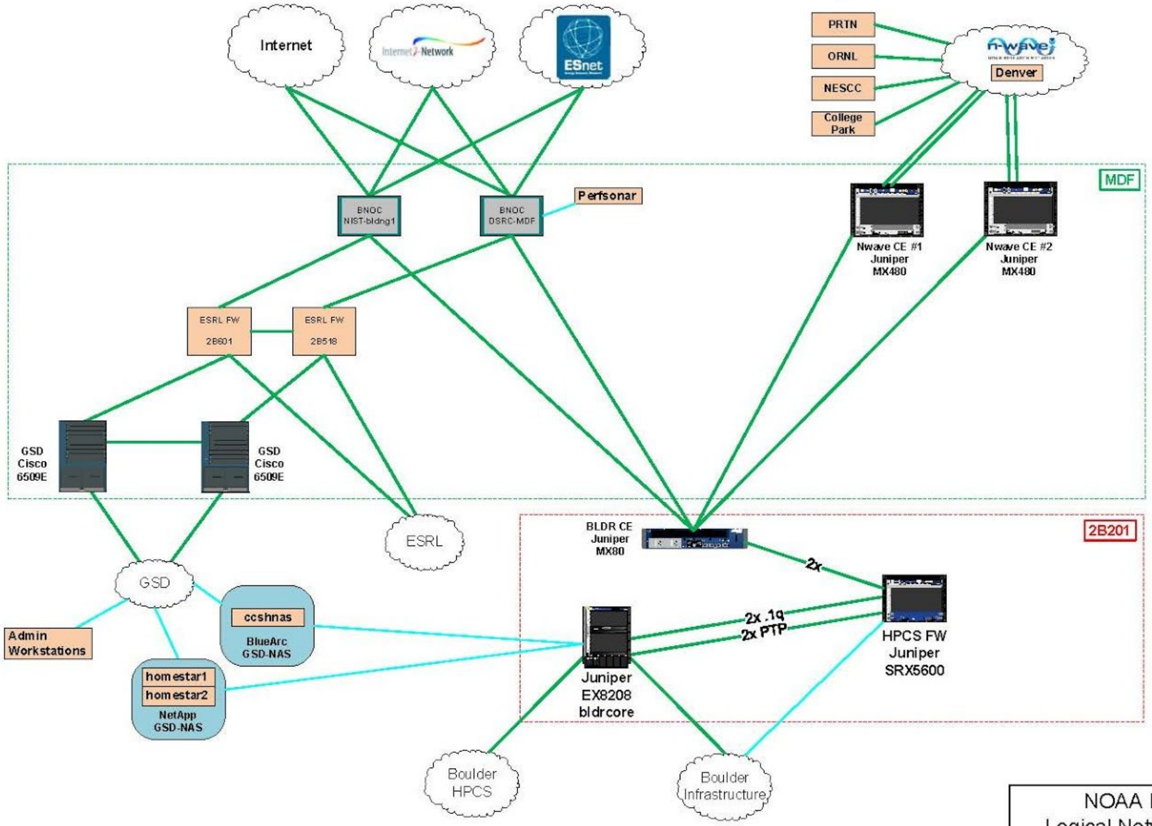
PROPERTY OF THE UNITED STATES GOVERNMENT
COPYING, DISSEMINATION, OR DISTRIBUTION OF THESE
DRAWINGS, PLANS, OR SPECIFICATIONS TO
UNAUTHORIZED USERS IS PROHIBITED
Do not remove this notice
Properly destroy documents when no longer needed

N

🟩	= xJet, HFIP p8
🟩	= tJet, HFIP p3b
🟩	= uJet, HFIP p4
🟪	= sJet, HFIP p5
🟩	= xJet, HFIP p7
□	= 2ft x 2ft Floor Tile



Boulder – External Connectivity



Legend

Ten Gigabit Ethernet	🟩
Gigabit Ethernet	🟦
10/100/1000TX Ethernet	■

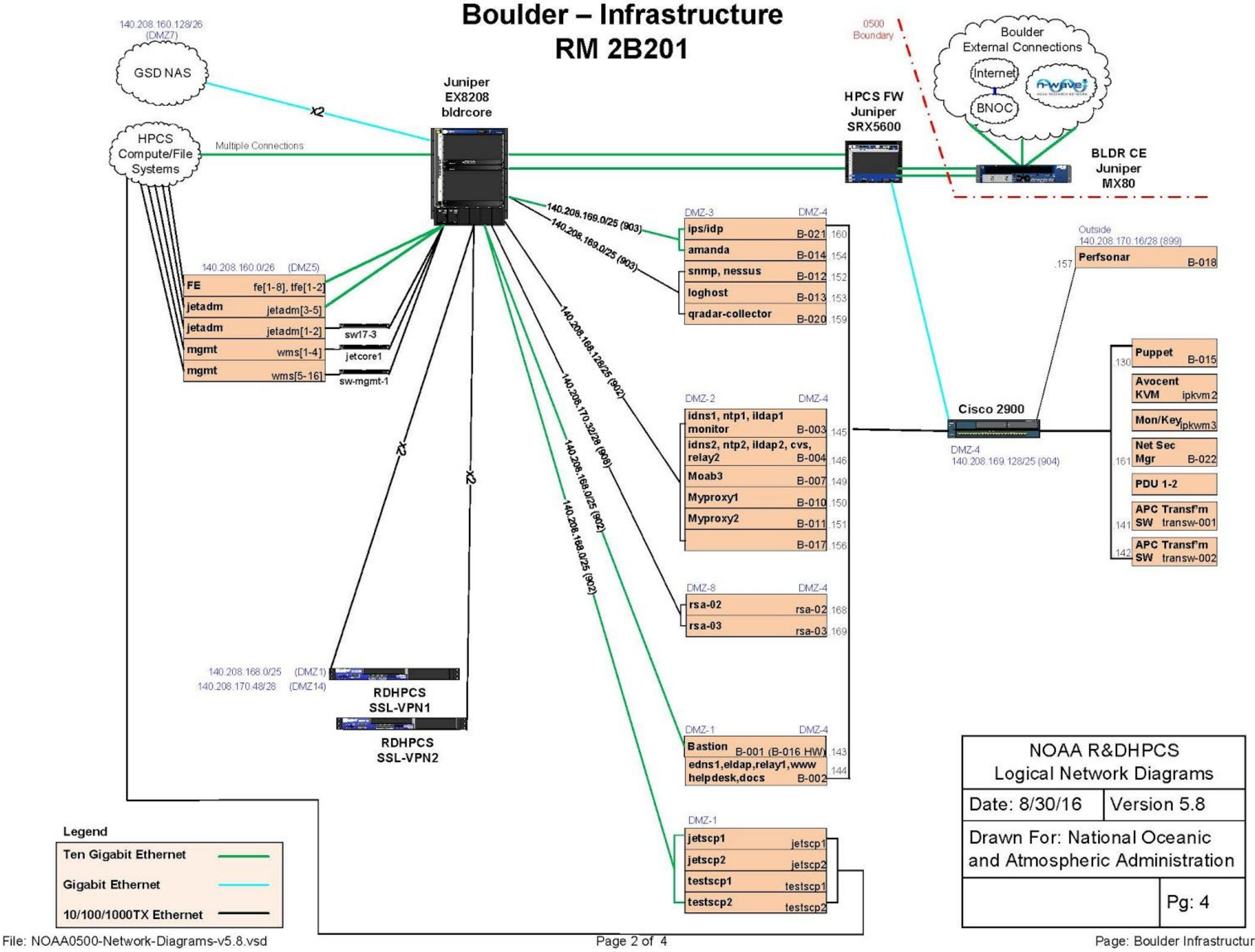
NOAA R&DHPCS
Logical Network Diagrams

Date: 10/31/13 | Version: 5.3

Drawn For: National Oceanic and Atmospheric Administration

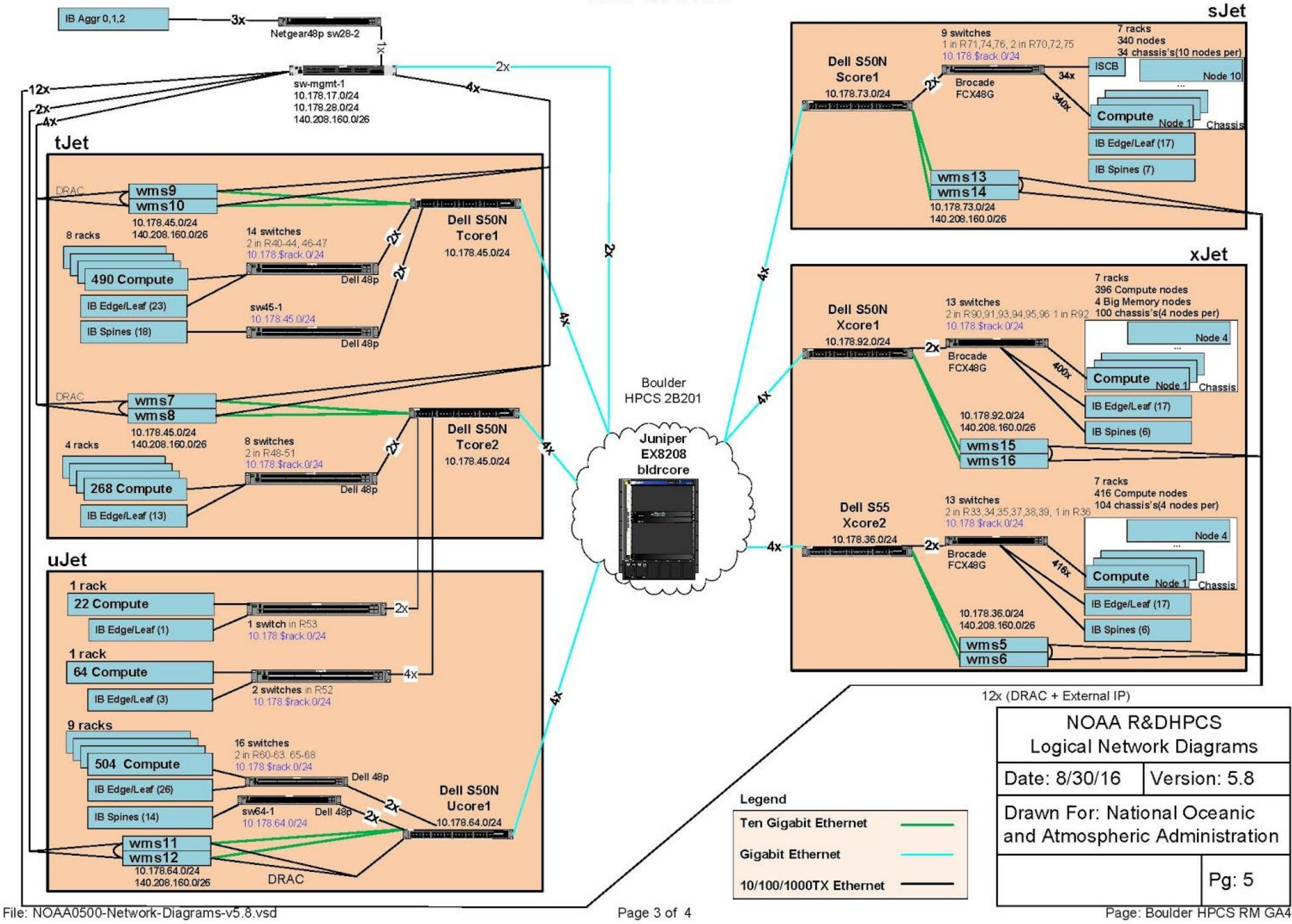
Pg: 3

Boulder – Infrastructure
RM 2B201



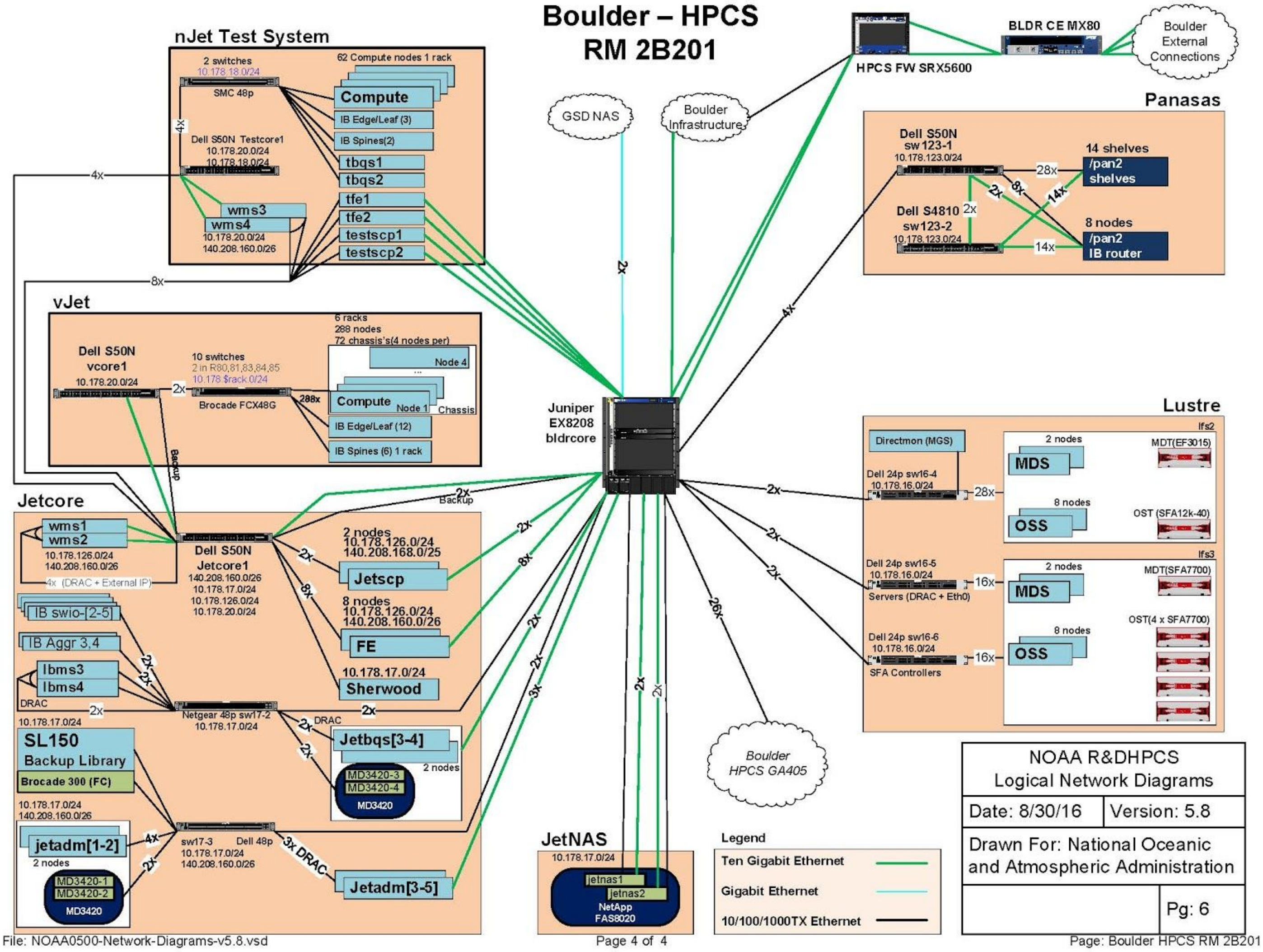
NOAA R&DHPCS Logical Network Diagrams	
Date: 8/30/16	Version 5.8
Drawn For: National Oceanic and Atmospheric Administration	
	Pg: 4

Boulder – HPCS
RM GA405



NOAA R&DHPCS Logical Network Diagrams	
Date: 8/30/16	Version: 5.8
Drawn For: National Oceanic and Atmospheric Administration	
	Pg: 5

Boulder – HPCS RM 2B201



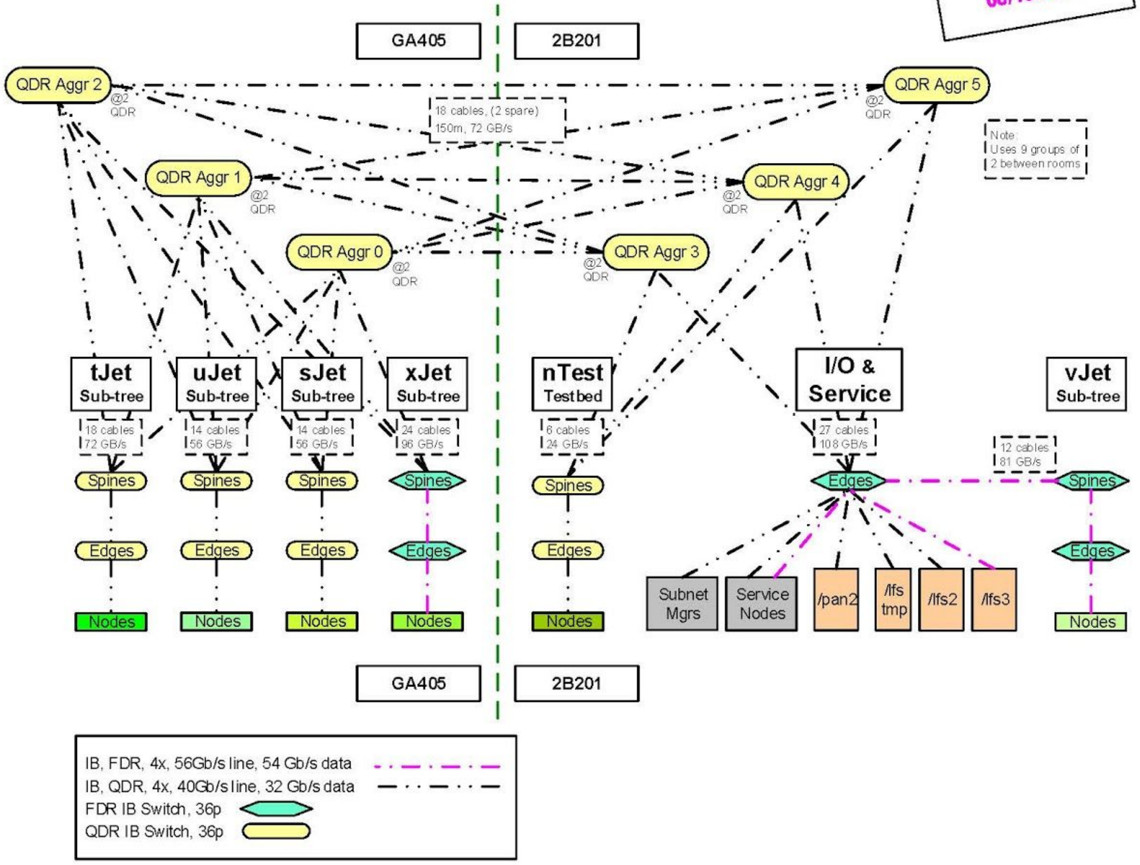
Legend

Ten Gigabit Ethernet	🟩
Gigabit Ethernet	🟦
10/100/1000TX Ethernet	⬛

NOAA R&DHPCS Logical Network Diagrams	
Date: 8/30/16	Version: 5.8
Drawn For: National Oceanic and Atmospheric Administration	
	Pg: 6

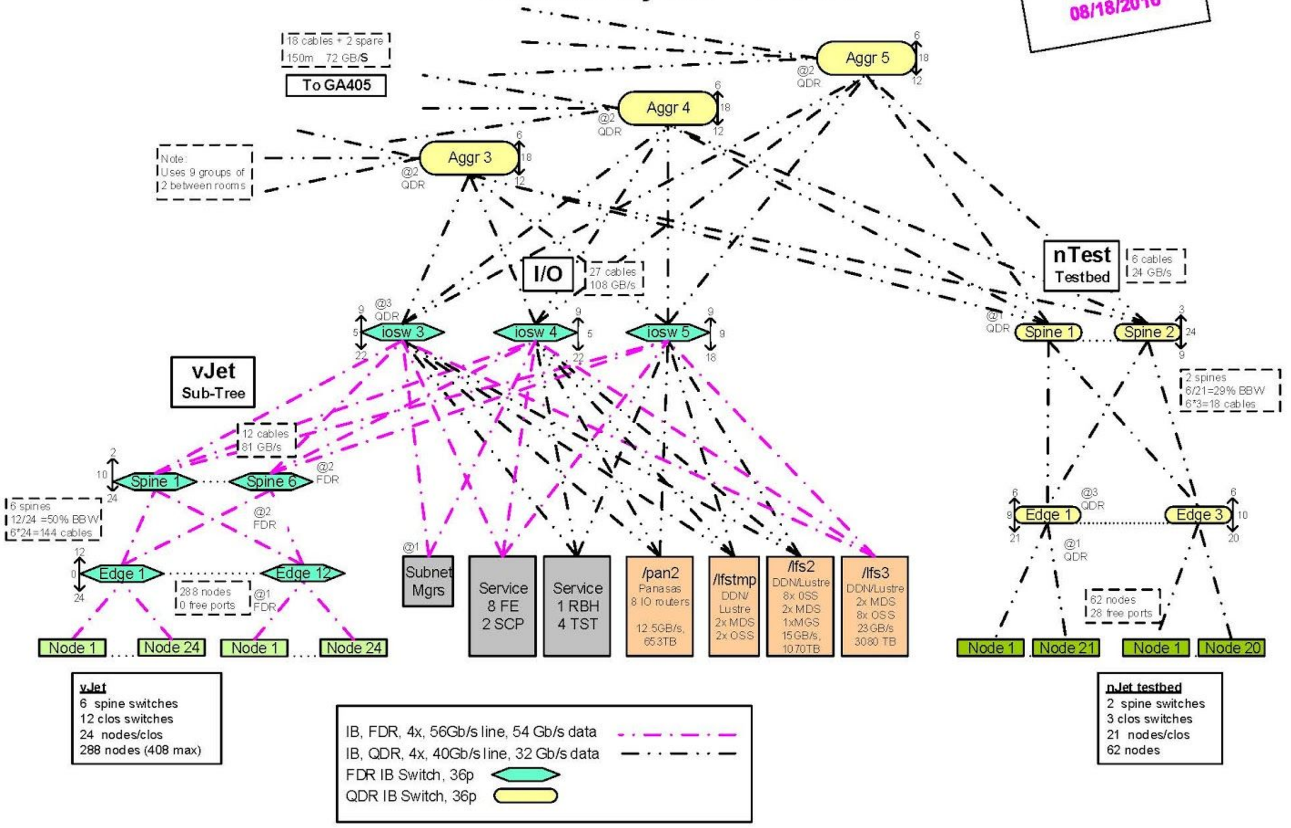
IB HSN HL Layout

HFIP p8
08/18/2016



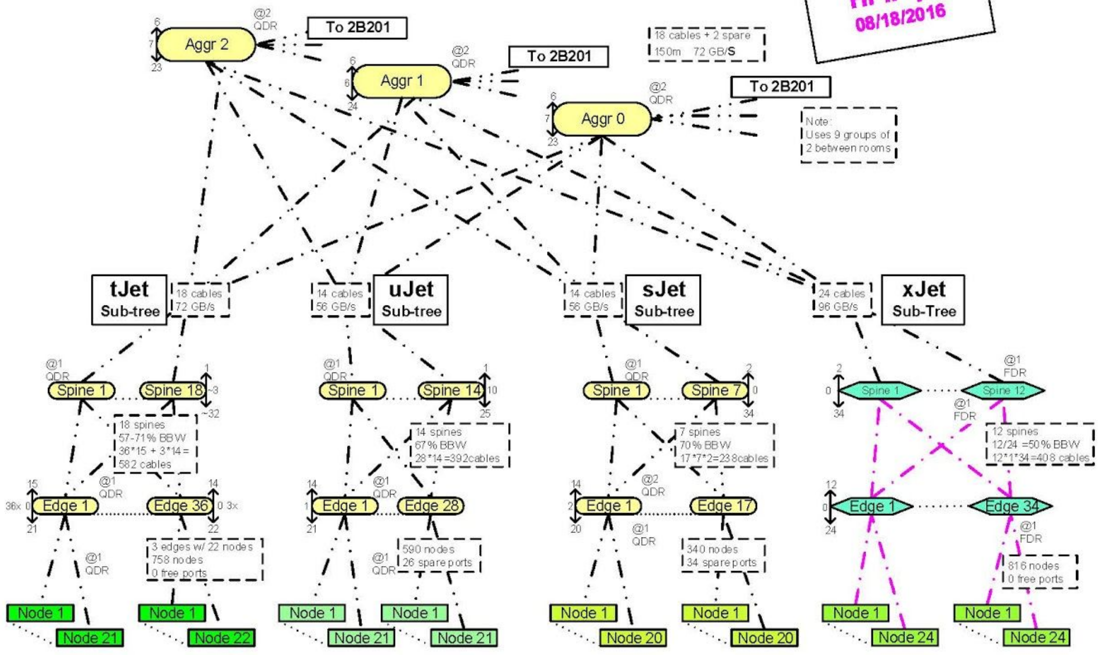
HFIP p8
08/18/2016

IB HSN Layout - 2B201



IB HSN Layout - GA405

HFIP p8
08/18/2016



tJet subtree
18 spine switches
36 edge switches (42 max)
21 or 22 nodes/edge
758 tJet nodes
(882 nodes max @ BBW)

uJet subtree
14 spine switches
28 edge switches (35 max)
21 nodes/edge
590 uJet nodes
(735 max @ BBW)

sJet subtree
7 spine switches
17 edge switches
20 nodes/clos
340 nodes (max @ BBW)

xJet subtree
12 spine switches
34 edge switches (34 max)
24 nodes/clos
816 nodes (816 max @ BBW)

IB, FDR, 4x, 56Gb/s line, 54 Gb/s data
IB, QDR, 4x, 40Gb/s line, 32 Gb/s data
FDR IB Switch, 36p
QDR IB Switch, 36p